中图法分类号:TP18;TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2026)08-3001-14

论文引用格式:Yang Y W, Ma J, Meng Y and Li K. 2026. PerRec: 3D dressed human reconstruction from perspective images via distortion feature decoupling and pseudo multi-view constraints. Journal of Image and Graphics, 31(8):3001-3014(杨源旺, 马健, 孟源, 李坤. 2026. 畸变特征解耦与伪多视角约束的透视图像穿衣人体三维重建. 中国图象图形学报, 31(8):3001-3014)[DOI:10.11834/jig.250413]

畸变特征解耦与伪多视角约束的透视图像穿衣人体三维重建

杨源旺,马健*,孟源,李坤*

天津大学智能与计算学部,天津　300354

摘　要:**目的**　单幅透视畸变图像的穿衣人体重建在虚拟试衣、数字人制作等应用中具有重要意义。然而,现有方法多采用正交或弱透视投影,忽略真实相机的成像规律,将深度简化为单一缩放因子,难以恢复透视畸变条件下真实的人体三维结构。为解决这一问题,提出一种面向透视图像的三维穿衣人体重建方法PerRec,通过畸变特征解耦与伪多视角约束缓解深度歧义与投影失真。**方法**　首先利用均匀分块与虚拟视角变换解耦图像位置与透视畸变;随后设计多尺度注意力增强网络以强化局部细节表达;并将SMPL网格(skinned multi-person linear model)转化为傅里叶占有率场以提升高频几何重建能力;最后通过伪多视角模块模拟多视角特征以增强对畸变的鲁棒性。**结果**　实验结果表明,PerRec在Thuman2.0和CustomHumans数据集上均取得显著提升,在倒角距离、法线一致性和点到表面距离等指标上相较现有方法平均提升60%～80%。**结论**　本文方法能够有效解决单幅透视图像下的结构失真问题,为真实场景中的高精度穿衣人体重建提供了可扩展的技术途径。

关键词:三维人体重建;透视畸变图像;单目相机;伪多视角;RGB图像

PerRec: 3D dressed human reconstruction from perspective images via distortion feature decoupling and pseudo multi-view constraints

Yang Yuanwang, Ma Jian*, Meng Yuan, Li Kun*

College of Intellgence and Computing, Tianjin University, Tianjin　300354, China

Abstract: **Objective** Reconstructing clothed human bodies from a single perspective-distorted image is a fundamental yet challenging problem in computer vision, with broad applications in virtual try-on, digital human creation, AR/VR content generation, and human-computer interaction. Although recent advances in monocular human reconstruction have achieved impressive progress, most existing approaches implicitly assume orthographic or weak-perspective projection models. These simplified assumptions neglect the actual imaging geometry of real cameras, reducing depth to a single global scaling factor and failing to capture the inherently nonlinear spatial variations caused by strong perspective distortion. Consequently, the reconstructed geometry often suffers from inaccurate limb proportions, inconsistent body scale, and distortion-dependent artifacts, particularly when the subject is close to the camera or occupies a non-central region of the image.

收稿日期:2025-09-05;**修回日期**:2025-12-29;**预印本日期**:2026-01-05

*通信作者:马健　jianma@tju.edu.cn; 李坤　lik@tju.edu.cn

基金项目:国家重点研发计划资助(2023YFC3082100);国家自然科学基金项目(62501416);天津市杰出青年科学基金项目(22JCJQJC00040);天津市自然科学基金项目(24JCYBJC01300)

Supported by: National Key R&D Program of China (2023YFC3082100); National Natural Science Foundation of China (62501416); Science Fund for Distinguished Young Scholars of Tianjin, China (22JCJQJC00040); Tianjin Municipal Natural Science Foundation (24JCYBJC01300)

Such limitations significantly undermine the applicability of current methods to real-world scenarios where perspective effects are inevitable. To address these shortcomings, This study proposes PerRec, a novel perspective-aware reconstruction framework specifically designed for recovering high-fidelity clothed human geometry from a single perspective-distorted RGB image. **Method** The key idea of PerRec is to explicitly account for perspective-induced depth ambiguity and projection nonlinearity through a combination of distortion decoupling, geometric field representation, and pseudo-multiple-view supervision. Unlike existing models that attempt to implicitly learn perspective from data, PerRec introduces structural priors and tailored network designs that directly target the characteristics of perspective distortion, thereby enabling a more accurate and physically plausible 3D inference. The proposed framework contains four major components. First, a distortion decoupling module is introduced to separate image position from the underlying perspective distortion pattern. The method effectively normalizes spatial variance while preserving layout-dependent cues by uniformly partitioning the input image and generating virtual viewpoint transformations for each region. This mechanism allows the model to effectively capture how projection distortion changes across the field of view, mitigating the tendency of conventional models to collapse depth variations into oversimplified estimates. Second, we design a multi-scale attention hourglass network (MA-HGNet) to robustly extract local and global features under varying degrees of distortion. The network integrates a multi-level hourglass structure with channel attention mechanisms to refine fine-grained patterns, such as clothing folds, body boundaries, and high-curvature regions that are particularly susceptible to distortion. MA-HGNet demonstrates enhanced sensitivity to spatially localized distortions while maintaining stable global shape perception compared with standard convolutional backbones. Third, we convert the SMPL mesh into a Fourier occupancy field (FOF) to effectively capture continuous geometry and enhance high-frequency surface details. This representation encodes occupancy values using Fourier basis functions, enabling the network to model sharp transitions, intricate clothing topology, and subtle geometric variations that are difficult to represent with parametric models alone. The proposed model unifies the advantages of parametric priors and implicit fields, resulting in substantially detailed and expressive reconstructions. Finally, considering that a single image inherently lacks sufficient viewpoint diversity, we introduce a pseudo-multi-view module that synthesizes multiple virtual observations during training. These synthesized views act as implicit geometric constraints, guiding the network to learn distortion-invariant features and improving its ability to reason about occluded or heavily distorted regions. This strategy significantly boosts reconstruction robustness without requiring additional input views or multi-camera setups. **Result** Extensive experiments are conducted on the Thuman2.0 and CustomHumans datasets, covering controlled and in-the-wild settings with diverse degrees of perspective distortion. Quantitative evaluations demonstrate that PerRec substantially outperforms state-of-the-art monocular reconstruction methods across multiple metrics, including chamfer distance, normal consistency, and point-to-surface error. Specifically, improvements of approximately 60%−80% are observed on average, highlighting the effectiveness of the proposed design in addressing global and local reconstruction quality. Qualitative comparisons further show that PerRec produces a stable body scale, highly realistic limb proportions, and detailed clothing geometry, particularly in strongly distorted images where competing methods typically fail. **Conclusion** The contributions of this study are threefold. First, this study provides a comprehensive analysis of the limitations of existing projection assumptions in monocular reconstruction and establishes perspective distortion as a core challenge in real-world 3D human modeling. Second, this study proposes a unified reconstruction framework that integrates distortion decoupling, implicit field representation, and pseudo-multi-view learning, offering a principled and effective solution for handling perspective effects. Third, this study demonstrates that accurate and high-fidelity clothed human reconstruction from a single perspective-distorted image is achievable without relying on multi-view capture systems or camera calibration. Overall, this study offers a practical and extensible perspective-aware solution for real-world human reconstruction, bridging the gap between theoretical projection modeling and data-driven 3D inference. The proposed approach not only enhances geometric accuracy under challenging perspective conditions but also expands the applicability of monocular reconstruction methods to everyday images captured by mobile devices and consumer cameras. Results indicate that perspective distortion, long regarded as a major obstacle for monocular reconstruction, can be effectively resolved by combining structural modeling cues with carefully designed learning-based strategies.

Key words: 3D human reconstruction; perspective-distorted image; monocular camera; pseudo multi-perspective; RGB image

0 引　言

单目图像的三维人体重建技术因其低成本、便捷性优势，成为计算机视觉与图形学的前沿研究方向（刘乐元 等，2024）。然而，在短视频创作和虚拟直播等场景中，近距离拍摄导致人体各部位在深度轴上产生显著差异，进而引发姿态扭曲和关节拓扑异常等问题。由于这种几何失真源于三维空间到二维映射的本质特性，传统方法通常依赖理想化的投影假设框架简化计算，但这会忽略透视畸变的影响，难以有效解耦透视畸变与真实人体结构的耦合关系，亟需通过建模透视投影的几何约束与人体解剖先验的协同优化机制实现算法层面的突破。

当前主流方法（Xiu 等，2023；Ho 等，2024；Tang 等，2026）普遍采用弱透视投影假设，模型通过线性近似显著降低优化复杂度，并规避了透视投影的非线性计算难题。然而，当直接应用于真实透视图像时，该假设会因忽略深度方向的非线性变化导致严重失真。如图 1 所示，在输入透视畸变图像时，现有的重建方法存在肢体比例失调和关键关节位姿偏移等明显缺陷。

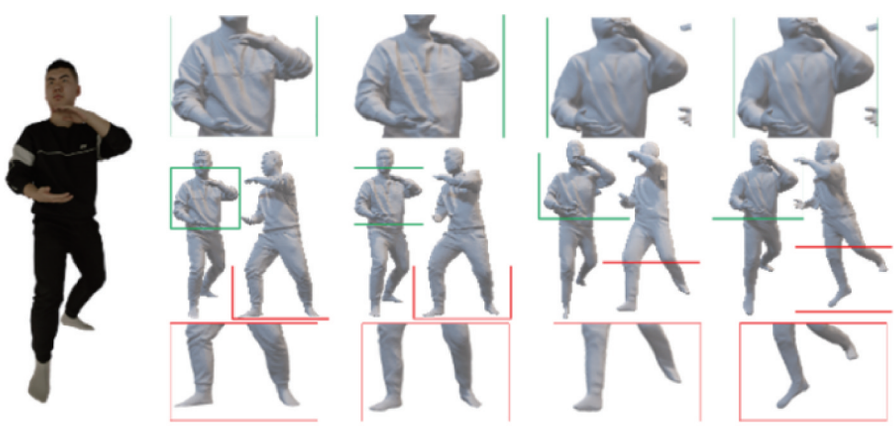

图 1　畸变图像的三维人体重建结果对比

Fig. 1　Comparison of 3D human body reconstruction results from distorted images

此外，当前透视畸变下的人体重建方法主要面向裸体人体建模（如 Zolly（Wang 等，2023）），仅输出参数化形体而缺乏服饰细节。这类方法虽然能适应透视投影的非线性畸变，但未考虑穿衣状态下的几何复杂性，导致细节重建能力不足。黄千芃等人（2024）提出一种单视角三维人体着装特征学习方法，通过分步学习姿态、褶皱与形状特征，结合柔性变形损失函数和有向距离场重建，获得高精度的重建，普骏程等人（2022）提出一种多阶段优化的三维人体重建方法，提取人体图像的语义、明暗和高频特征，再基于局部深度特征构建有向距离场隐式表征三维几何，通过着装层次损失函数优化生成粗糙模型，最后融合明暗和高频特征在 UV 空间定位并细化服装褶皱，得到高精度三维着装人体模型。上述两个方法聚焦于着装人体的几何细节重建，一定程度上弥补了裸体建模方法的局限性，但在透视投影畸变与单目图像信息解耦方面处理能力不足。最后，单图像缺少有效信息以及多视角几何约束，难以区分投影畸变（相机位姿引起）与真实形变（动作或布料物性引起）。例如，肘部弯曲褶皱和俯仰角导致的透视缩短在二维图像中可能呈现相似位移模式，但单目观测无法解耦二者。故透视图像下的穿衣人体重建的挑战可总结为透视投影导致的非线性畸变、单目图像有效信息缺失以及服饰几何细节重建不足。

为了克服上述问题，本文提出一种新的三维穿衣人体重建方法——畸变特征解耦与伪多视角约束的透视图像三维穿衣人体重建方法 PerRec。该方法通过透视畸变特征自适应和伪多视角约束两大模块，有效解耦了投影畸变与图像位置的关系，弥补了单视角图像在几何细节恢复上的不足，从而实现从单幅透视图像中高精度重建三维穿衣人体。实验结果显示，相比于目前最新方法，所提方法在 Thuman2.0（Yu 等，2021）和 CustomHumans（Ho 等，

2023)数据集上的定量和定性均达到最优,在倒角距离(chamfer distance, CD)、法线一致性(normal consistency, NC)和点到表面距离(point-to-surface, P-to-S)3个指标上比目前最优方法分别提高82.9%、62.0%、76.16%和72.3%、84.7%、80.2%。

本文的创新点如下:1)提出一种畸变特征解耦与伪多视角约束的透视图像三维穿衣人体重建方法PerRec,旨在解决透视图像下的穿衣人体三维重建问题。2)提出透视畸变特征自适应模块,解决透视投影导致的非线性畸变。通过均匀分块和虚拟视角变换,解耦图像位置与畸变效应;并设计多尺度注意力增强网络,实现局部特征的精细化建模。3)提出伪多视角模块,通过虚拟光心移动策略模拟差异化畸变,并设计权重分配机制以模拟多视角几何约束,从而解决单图像信息缺失问题。4)本研究在多个数据集均取得最优性能,相较于现有最先进方法,在倒角距离、法线一致性以及点到表面距离3个关键指标上均实现了60%以上的显著提升。

1 相关研究

1.1 单图像三维穿衣人体重建

近年来,从单幅图像中重建3D人体的研究(Xiu等,2022,2023)逐渐成为热点。根据重建方法的不同,现有研究主要分为显式重建和隐式重建两大类。

基于显式表面的三维重建通过参数化人体模型(如SMPL(skinned multi-person linear model)(Loper等,2015)、SMPL-X(SMPL expressive)(Pavlakos等,2019)、STAR(sparse trained articulated human body regressor)(Osman等,2020))直接定义表面几何的拓扑结构与空间坐标,将人体形状与姿态解耦为低维可解释参数。例如,ARCH(animatable reconstruction of clothed humans)(Huang等,2020)通过引入层次化网络结构,将人体姿态参数与衣物形变参数联合优化,通过分层建模策略,将人体姿态、形状与衣物形变解耦为多个子任务,分别优化后再进行全局融合。SCANimate(skinned clothed avatar network)(Saito等,2021)则通过结合动态捕捉数据与物理仿真技术,在预测顶点偏移时融入了布料动力学特性,显著提升了衣物的自然度。基于隐式表面的三维重建技术通过连续函数隐式定义人体表面几何,突破了传统显式建模的拓扑限制,能够直接表征任意复杂度的穿衣人体形态。以PIFuHD(multi-level pixel-aligned implicit function for high-resolution 3D human digitization)(Saito等,2020)为代表的像素对齐隐式函数方法,通过多尺度特征融合机制将高分辨率图像映射至三维空间,实现了高精度人体重建,其关键在于设计像素级对齐的隐式场编码器,使得每个三维点能够关联到图像局部区域的语义与几何特征。然而,纯隐式方法在缺乏显式人体先验时易产生肢体拓扑错误(如手指粘连或足部穿透),为此,ICON(implicit clothed humans obtained from normals)(Xiu等,2022)提出将参数化人体模型约束与隐式场联合优化,通过显式模板驱动隐式表面变形,既保留了隐式建模的灵活性,又规避了肢体断裂等非物理合理现象,尤其在处理宽松衣物时能有效维持人体关节的运动学一致性。ECON(explicit clothed humans optimized via normal integration)(Xiu等,2023)通过融合显式参数化模型与隐式细节补全网络,在保持主体结构合理性的同时恢复了衣物局部形变。尽管两类方法在三维人体建模中取得了显著进展,其性能仍受限于实际成像过程中的几何失真问题,尤其是由相机透视投影引起的畸变效应。

1.2 透视畸变下的三维重建

透视畸变是由于相机成像模型中的透视投影特性导致的几何形变现象,在人体网格重建中,透视畸变的单目三维人体重建会导致比例失真、关节位置偏差等问题。当前解决透视畸变的技术路线可分为两类:基于几何标定的显式修正方法和基于深度学习的自适应修正方法。几何标定方法通过精确估计相机参数显式校正透视畸变,其研究主要集中在相机标定算法与投影模型优化上。Heikkila和Silven(1997)通过引入非线性优化技术提升标定精度,特别是在大畸变场景下的鲁棒性。张正友标定法(Zhang,2000)通过棋盘格角点的像素坐标与物理坐标计算相机内参矩阵与畸变参数,为后续研究奠定了理论基础。但其依赖标定板与场景假设的局限性限制了其在单目图像人体重建中的应用。深度学习技术为透视畸变校正提供了新的解决思路,通过数据驱动的方式隐式学习相机参数与人体几何的映射关系。Zolly(Wang等,2023)提出了动态焦距优化模块,通过分析人体解剖学比例反推焦距,显著减小了重建比例误差。该方法通过引入人体比例先验,能够在不依赖标定板的情况下自适应调整焦距参数。

CLIFF(carrying location information in full frames)(Li等,2022)加上额外的输入信息使得神经网络自发地考虑人和相机的水平角度。NeuMan(Jiang等,2022)利用 COLMAP(structure-from-motion and multi-view stereo)估计的相机参数和2D-3D关节对应点,通过透视n点算法解决人体姿态估计器与神经辐射场之间的空间对齐问题,同时"至少有一帧人物站立在地面"的假设确保了人体与场景交互的物理合理性以及对新视角的适应性。

现有针对畸变图像的三维人体重建研究多集中于裸模,本质上假设人体表面为刚性几何,未考虑衣物形变等非刚性因素。然而,真实场景中的人体通常被衣物覆盖,这使得透视畸变与衣物形变在成像过程中呈现非线性耦合效应。这种耦合使得传统基于裸模的线性解耦假设在穿衣人体场景中失效,进而引发姿态扭曲、衣物褶皱等问题。因此,本文提出一种畸变特征解耦与伪多视角约束的透视图像三维穿衣人体重建方法PerRec,旨在解决先前工作在透视畸变图像重建任务中精度低等问题。

2 本文方法

本文提出畸变特征解耦与伪多视角约束的透视图像三维穿衣人体重建方法PerRec,以实现透视畸变图像下的高精度穿衣人体重建,整体框架如图2所示。提出透视畸变特征自适应,通过均匀分块和虚拟视角变换解耦图像位置与非线性畸变效应,结合多尺度注意力网络实现局部特征精细化建模;同时设计伪多视角模块,利用虚拟光心偏移模拟差异化畸变,并通过权重分配机制模拟多视角几何约束,有效解决单目图像信息缺失问题。

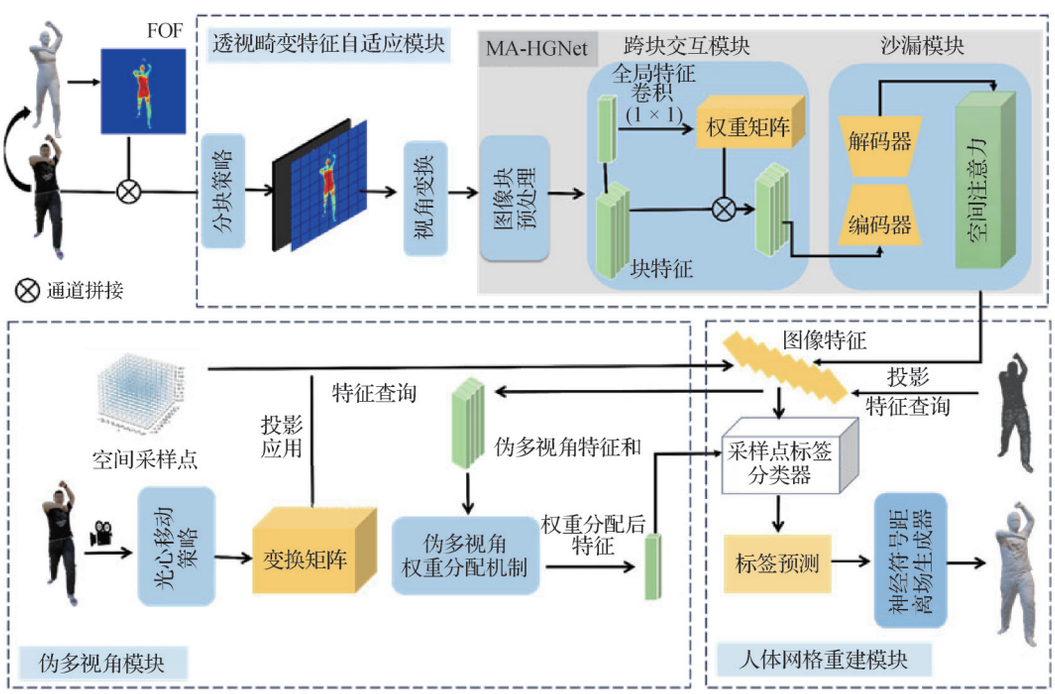

图2 PerRec方法框架示意图

Fig. 2 Pipeline of the PerRec

2.1 总览

针对人体三维几何的高效建模需求,PerRec使用人体参数模型,通过傅里叶级数实现三维占据场的轻量化二维表征(Feng等,2024)。为消除人体姿态与尺度的差异性,将SMPL网格顶点从原始模型空间线性映射至标准化立方体空间,具体为

$$\boldsymbol{V}_{\mathrm{NDC}} = 2 \cdot \frac{\boldsymbol{V}_{\mathrm{mesh}} - \boldsymbol{V}_{\mathrm{min}}}{\boldsymbol{V}_{\mathrm{max}} - \boldsymbol{V}_{\mathrm{min}}} - 1 \tag{1}$$

式中,$(\boldsymbol{V}_{\mathrm{min}}, \boldsymbol{V}_{\mathrm{max}})$为网格顶点的坐标极值,$\boldsymbol{V}_{\mathrm{mesh}}$为网格点位置坐标。映射后所有顶点位于$[-1,1]^3$空间。随后,通过投影将三维面片映射至二维网格平面,同步记录每个像素单元$p(u,v)$对应的深度分布$\{z_i\}_{\{i=1\}}^{N}$(N为投影面片数)。遵循FOF(Fourier occu-

pancy field)原始论文设置(Feng等,2022),沿深度方向(z轴)对每个像素的占用函数进行傅里叶级数积分,计算前32阶系数构建频域特征;最后,整合空间—频域信息,输出分辨率为512×512像素、通道维度为32的多通道图像$\boldsymbol{S}_{\mathrm{FOF}}$,实现三维人体几何的轻量化二维表征。通过SMPL参数约束,保证不同姿态下频域特征的拓扑连续性。

随后,将$\boldsymbol{S}_{\mathrm{FOF}}$与输入图像拼接在一起,将拼接后的图像经过透视畸变特征自适应模块,使用3D采样点经过投影对图像特征进行插值提取,随后利用隐式函数预测空间点的符号距离值,结合均方根传播(root mean square propagation, RMSProp)优化器训练,最终生成高精度的人体重建模型。

在测试阶段引入伪多视角模块,突破单视角的固有局限。所提方法首先对图像进行仿射变换,生成4个虚拟相机的内参矩阵,每个相机聚焦于512×512像素图像的不同子区域中心,模拟多视角观测。将虚拟相机内参用于采样点的投影,生成4种不同视角的2D投影点,对投影点运用权重分配机制,距离光心越近的权重越大,往外逐步减小,保证权重总和为1。将权重乘以该位置对应的特征,得到基于多视角的总特征,用于人体重建。

2.2 透视畸变特征自适应模块

透视图像特征自适应是将分块作为基础条件,通过视角变换进行透视畸变处理,再依据MA-HGNet模型提取图像特征。

2.2.1 分块策略

在传统的成像过程中,由于透视投影存在的畸变,物体在图像中形状和大小会随着拍摄位置的变化而变化,故这种畸变会导致以往模型在处理图像时,难以准确理解三维结构。为此,本文方法以图像块为单位进行处理,以解耦图像畸变与图像位置的相互依赖关系。给定一幅尺寸为$H\times W$的RGBA图像,PerRec首先将其均匀划分为$p\times p$个不重叠的子块。每个子块的大小为$p_{ij}\in\mathbf{R}^{\frac{H}{p}\times\frac{W}{p}\times 4}$,同时确定每个子块的中心点在原始图像中的坐标位置,即光心位置(x_{ij},y_{ij})为

$$x_{ij}=\frac{W}{p}\times\left(j+\frac{1}{2}\right) \tag{2}$$

$$y_{ij}=\frac{H}{p}\times\left(i+\frac{1}{2}\right) \tag{3}$$

式中,W和H分别表示图像的高和宽;i和j分别表示子块的行索引和列索引,取值范围为0,1,2,…,p-1。

2.2.2 视角变换

首先,定义一个虚拟相机,通过构建仿射变换矩阵,生成与真实相机对应的虚拟相机内部参数,裁剪区域的某点$\hat{p}$都可以用原始图像$\boldsymbol{p}$的位置经过仿射变换得到,该仿射变换可写为

$$\hat{p}=\boldsymbol{C}\boldsymbol{p},\boldsymbol{C}=\begin{bmatrix}S_x & C_x & A_x\\ C_y & S_y & A_y\\ 0 & 0 & 1\end{bmatrix} \tag{4}$$

式中,$\boldsymbol{A}=[A_x,A_y]$表示二维平移,$\boldsymbol{S}=[S_x,S_y]$表示在两个方向上的缩放因子,$\boldsymbol{C}=[C_x,C_y]$表示倾斜因子。虚拟相机的内参为$\boldsymbol{K}^{\mathrm{virt}}=\boldsymbol{C}\boldsymbol{K}$。其中,$\boldsymbol{K}$为原始相机的内参矩阵。

其次,利用虚拟相机内部参数对图像块进行重投影变换,根据虚拟相机中心相对于原始图像中心的偏移量计算得到相机旋转矩阵,具体计算为

$$\boldsymbol{R}_{\mathrm{virt}\to\mathrm{real}}=\begin{bmatrix}\dfrac{1}{\sqrt{1+p_x^2}} & \dfrac{-p_xp_y}{\sqrt{(1+p_x^2+p_y^2)(1+p_x^2)}} & \dfrac{p_x}{\sqrt{1+p_x^2+p_y^2}}\\ 0 & \dfrac{\sqrt{1+p_x^2}}{\sqrt{1+p_x^2+p_y^2}} & \dfrac{p_y}{\sqrt{1+p_x^2+p_y^2}}\\ \dfrac{-p_x}{\sqrt{1+p_x^2}} & \dfrac{-p_y}{\sqrt{(1+p_x^2+p_y^2)(1+p_x^2)}} & \dfrac{1}{\sqrt{1+p_x^2+p_y^2}}\end{bmatrix} \tag{5}$$

式中,p_x和p_y分别为图像块中心点位置与原始图像中心点位置的横纵轴距离。

原始图像到裁剪区域的映射用下式实现,具体为

$$\tau(u,v,s,K)=\boldsymbol{K}^{\mathrm{virt}}R_{\mathrm{virt}\to\mathrm{real}}^{-1}\boldsymbol{K}^{-1} \tag{6}$$

式中,$\boldsymbol{K}$是真实相机的内参矩阵,$\boldsymbol{K}^{\mathrm{virt}}$是虚拟相机的内参矩阵,$\boldsymbol{R}_{\mathrm{virt}}$是真实相机中心到虚拟相机某视角下的旋转矩阵。

经过上述虚拟相机映射后,得到了畸变效果一致的图像块$\boldsymbol{P}_i\in\mathbf{R}^{\frac{H}{p}\times\frac{W}{p}\times 3}$,使得在该视角下物体的形状和大小不再受到透视畸变的影响。

值得注意的是,本文方法的视角变换并非旨在恢复真实相机模型,也不是用于估计诸如焦距或主点位置等物理内在参数。相反,本文的表述采用相对透视表示,透视畸变程度通过每个图像块内虚光中心的归一化偏移来建模。给定分辨率为$W\times H$的输入图像,每个块区域被赋予一个虚拟主点,该点

(c_x', c_y')纯粹是根据其相对空间位置计算，而非任何物理相机校准。这一机制反映了一个常见的观察点：远离全球图像中心的区域会经历更强的透视效应。因此，模型学习了位置相关的透视调制，有效将图像外观与空间变化的畸变耦合，无需焦距、畸变系数或完整的内在矩阵。本文方法提供了通用且无需校准的透视变化近似，专为穿着衣服的人形几何重建量身定制。这种设计使PerRec无需任何摄像机元数据即可运行，并确保对各种真实视角畸变的鲁棒性。

2.2.3　多尺度注意力增强网络

MA-HGNet模型架构主要由图像块预处理、跨块交互模块和沙漏模块3个部分组成，其中，图像块预处理首先采用（3×3）卷积核配合批归一化操作对输入图像进行初始特征变换，生成具有64个通道的基础特征表示。

接着，为强化局部细节建模并实现全局上下文感知，本模块创新性地设计了基于通道注意力的特征增强机制，其核心思想是通过细粒度局部特征分析与跨块交互实现自适应特征校准。具体而言，首先将经过图像块预处理后的特征采用跨块特征聚合策略计算全局平均特征$\boldsymbol{F}_{\text{avg}}$，该操作通过对所有图像块的卷积特征$\boldsymbol{F}_{\text{conv}}(\boldsymbol{P}_i)$进行算术平均，构建具有全局统计意义的特征基准。具体为

$$\boldsymbol{F}_{\text{avg}} = \frac{1}{N_p}\sum_{i=1}^{N_p}\boldsymbol{F}_{\text{conv}}(\boldsymbol{P}_i) \tag{7}$$

式中，N_p为图像块数量，$\boldsymbol{F}_{\text{conv}}$为经过卷积操作处理后的图像特征。为进一步挖掘块间关联性，模块引入可学习的1×1卷积层建立跨块特征交互通道，该层通过分析各通道在全局上下文中的贡献度，动态生成通道重要性权重矩阵为

$$\boldsymbol{a}_i = \sigma(Conv_{1\times1}([\boldsymbol{F}_{\text{conv}}(\boldsymbol{P}_i)\|\cdots\|\boldsymbol{F}_{\text{conv}}(\boldsymbol{P}_i)])) \tag{8}$$

通过学习到的通道重要性权重矩阵$\boldsymbol{a}_i$与原始块特征$\boldsymbol{F}_{\text{conv}}(\boldsymbol{P}_i)$进行通道级点乘运算，并与全局基准特征$\boldsymbol{F}_{\text{avg}}$进行加权融合，实现局部特征的自适应增强，该机制使得网络在特征融合过程中自动平衡局部细节与全局语义的贡献比例，显著提升模型对多尺度特征的表达能力。

$$\boldsymbol{F}_i' = a_i \odot \boldsymbol{F}_{\text{conv}}(\boldsymbol{P}_i) + \beta \cdot \boldsymbol{F}_{\text{avg}} \tag{9}$$

式中，⊙表示矩阵点乘，β为可学习参数。

最后，为了进一步提升网络对复杂特征的建模能力，参考PIFu（Saito等，2019），采用4个沙漏模型，其中，前3个沙漏模型的输出均通过（1×1）的卷积和$\boldsymbol{F}_i'$融合保留不同层次语义特征，每个沙漏模块的输出特征都会作为下一个沙漏模块的输入进行递进式特征优化，通过这种迭代精修机制不断修正特征表示中的几何细节和语义信息，最终输出的多层次融合特征既包含精细的局部结构描述，又整合了全局理解，为后续的三维重建任务提供了鲁棒的特征表示。

2.3　伪多视角模块

针对单幅图像成像视角受限，缺乏多视角几何信息，导致不同的透视距离可能在二维图像中呈现相似位移。然而现有的方法，例如ICON（Xiu等，2022）、ECON（Xiu等，2023）等，通常采用弱透视模型，忽视了深度问题，对畸变图像并不适应。因此，提出伪多视角模块，通过光心移动策略和权重分配机制，增强网络对畸变图像的鲁棒性。

本文通过在成像平面上设置一定个数的虚拟相机主点坐标，有效模拟了不同视角下的成像几何变化。这种设计使得每个虚拟相机的光心相对于原始图像中心产生不同程度的偏移，通过透视投影变换自然地引入不同程度的图像畸变效应。该方法不仅能够从单幅输入图像高效地模拟多视角、多畸变程度的数据，还能保持图像间的几何一致性，为后续的特征提取提供了更加丰富且精确的几何约束信息。

2.3.1　光心移动策略

为了模拟多视角观测并增强模型对视角变化的鲁棒性，PerRec提出了一种基于虚拟光心平移的数据增强方法。具体而言，给定原始输入图像及其对应的相机内参矩阵：$\boldsymbol{K}_0 = \begin{bmatrix} f & 0 & c_x^{(0)} \\ 0 & f & c_y^{(0)} \\ 0 & 0 & 1 \end{bmatrix}$，$f$为焦距，$(c_x^{(0)}, c_y^{(0)})$为主点坐标。为模拟光心平移，定义虚拟相机的内参矩阵为

$$\boldsymbol{K}_i = \begin{bmatrix} f & 0 & c_x^{(i)} \\ 0 & f & c_y^{(i)} \\ 0 & 0 & 1 \end{bmatrix} \tag{10}$$

式中，$i = 1, \cdots, N$，主点偏移量$(c_x^{(i)}, c_y^{(i)})$按以下规则生成：首先，将图像平面均匀划分为$M \times M$个子区域，每个子区域边长为

$$s = \left\lfloor \frac{W}{2M} \right\rfloor \tag{11}$$

式中，W为图像宽度。

然后，本文以各子区域的中心坐标作为虚拟主点位置，构造$N = M^2$个虚拟相机的内参矩阵，此操作等效于在成像平面坐标系下沿x和y方向平移相机光心，从而生成N个不同的虚拟视角，即

$$\begin{cases} c_x^{(i)} = (2m - 1)s \\ c_y^{(i)} = (2n - 1)s \end{cases}$$
$$m, n = 1, \cdots, M \tag{12}$$

式中，$c_x^{(i)}$和$c_y^{(i)}$表示虚拟相机内参中心点横纵坐标。为了平衡修正效果和计算效率，本文设置4个对称分布的虚拟相机主点坐标(128, 128)、(128, 384)、(384, 384)和(384, 128)。通过这种可控的虚拟视角生成机制，PerRec能够在数据层面增强模型对于不同视角和畸变情况的适应能力，从而提升在真实复杂成像条件下的特征表示鲁棒性。

2.3.2　伪多视角权重分配机制

为有效整合虚拟光心扰动生成的伪多视角特征，PerRec提出一种基于几何感知的自适应权重分配策略。该机制通过分析3D投影点在多视角图像平面上的几何分布特性，动态计算各视角特征的贡献权重，实现鲁棒的特征融合。首先根据投影点到图像中心的距离，计算初始权重：

给定3D点云$\boldsymbol{P} \in \mathbf{R}^{B \times 3 \times N}$($B$为批次大小，$N$为采样点数)，通过投影变换将其映射至各伪视角的归一化图像坐标系，计算式为

$$x_{ij} = \boldsymbol{T}_{\text{proj}}(\boldsymbol{P}, \boldsymbol{K}_j) \tag{13}$$

式中，$j = 1, \cdots, M$，$\boldsymbol{K}_j$为第j个虚拟视角的内参矩阵，$\boldsymbol{T}_{\text{proj}}$为可微投影算子。$x_{ij}$为2D投影点。定义可见性掩码$m_j \in \{0, 1\}^{B \times N}$为

$$m_j = \boldsymbol{I}(\left| x_{ij}^x \leqslant 1 \right|) \odot \boldsymbol{I}(x_{ij}^y \leqslant 1) \tag{14}$$

式中，$\boldsymbol{I}(\cdot)$为指示函数，有效投影区域为$[-1, 1]^2$；x_{ij}^x和x_{ij}^y表示3D投影点的横纵坐标。

为平衡不同视角的贡献，采用基于径向距离的权重函数为

$$w_j^{\text{raw}} = \frac{1}{\sqrt{(x_{ij}^x)^2 + (x_{ij}^y)^2 + \epsilon}} \tag{15}$$

式中，$\epsilon = 10^{-9}$用于数值稳定性。该权重赋予靠近图像中心的投影点更高置信度(中心区域畸变较小)，同时抑制边缘像素的噪声干扰。通过可见性掩码修正后，最终权重为

$$w_j = \frac{m_j \odot w_j^{\text{raw}}}{\sum_{k=1}^{M} m_k \odot w_k^{\text{raw}}} \tag{16}$$

归一化操作确保各视角权重和为1，实现自适应加权融合。对每个3D投影点$\boldsymbol{P}_i$，从各伪视角特征图$\boldsymbol{F}_j \in \mathbf{R}^{B \times C \times H \times W}$中提取对应位置的特征向量为

$$\boldsymbol{f}_j = BilinearSample(\boldsymbol{F}_j, x_{ij}) \tag{17}$$

式中，$BilinearSample$为双线性采样。然后，基于权重w_j进行通道级加权融合，具体为

$$\boldsymbol{f}_{\text{fused}} = \sum_{j=1}^{M} w_j \cdot \boldsymbol{f}_j \tag{18}$$

该操作通过可微双线性插值实现，保留空间连续性。

2.4　损失函数

给定一个训练样本$\boldsymbol{I}$，损失函数L可以表示为

$$L = \frac{1}{N} \sum_{k=1}^{N} \left\| \varphi\left(\boldsymbol{F}_{\text{cnn}}\left(\boldsymbol{I}\left(\boldsymbol{\pi}\left(\boldsymbol{X}_k \right) \right) \right), z\left(\boldsymbol{X}_k \right) \right) - \varphi^{\text{gt}}(\boldsymbol{X}_k) \right\| \tag{19}$$

式中，$\boldsymbol{X}_k \in \mathbf{R}^3$为第$k$个三维空间采样点坐标。$\boldsymbol{\pi}$: $\mathbf{R}^3 \to \mathbf{R}^2$为相机投影矩阵，$\boldsymbol{F}_{\text{cnn}}\left(\boldsymbol{I}\left(\boldsymbol{\pi}\left(\boldsymbol{X}_k \right) \right) \right)$为图像$\boldsymbol{I}$的$\boldsymbol{\pi}\left(\boldsymbol{X}_k \right)$位置提取的特征。$z(\boldsymbol{X}_k)$为三维点$\boldsymbol{X}_k$的深度编码特征；$\varphi(\cdot)$为隐式函数，输出符号距离场(signed distance field, SDF)；N为空间采样点总数。计算式可分解为3个计算阶段：计算投影坐标、特征融合和场值预测。

3　实验与结果分析

3.1　数据集与实验配置

目前，大部分单图像穿衣人体重建方法采用正交/弱透视图像，故本文方法使用THuman2.0(Yu等，2021)与CustomHumans(Ho等，2023)提供的人体扫描模型渲染透视图像，训练集包含381个Thuman2.0模型(覆盖BMI(body mass index)指数18.5～28.9的多样化体型)，测试集由145个Thuman2.0模型与全部CustomHumans数据集模型(647)组成，每个模型渲染256幅图(沿y轴均匀采样)，所有图像均采用透视相机和预计算的辐射传输渲染器进行渲染，图像分辨率为512×512像素。为了获得相应的SMPL模型，使用OpenPose在渲染的图像上检测二维关节点，并使用SMPLify-X算法的多视图修改版生成SMPL模型。

本文算法架构基于 PyTorch 实现，并在单个 NVIDIA RTX 3090 GPU 上进行训练。优化器采用 RMSprop，批大小为 4，初始学习率为 1×10^{-3}，权重衰减为 10^{-1}。模型共训练 15 个轮次，在单个 NVIDIA RTX 3090 上完成训练大约需要 4 d。

3.2 度量指标

为评估单图像人体重建的准确性，本文方法采用了广泛使用的评价指标与现有方法进行定量评估，包括倒角距离、法线一致性和点到表面的距离，单位均为 cm。

1）倒角距离的定义为

$$CD(\boldsymbol{P},\boldsymbol{Q}) = \frac{1}{|\boldsymbol{P}|}\sum_{p\in P}\min_{q\in Q}\left\|\boldsymbol{p}-\boldsymbol{q}\right\|^2 + \frac{1}{|\boldsymbol{Q}|}\sum_{q\in Q}\min_{p\in P}\left\|\boldsymbol{q}-\boldsymbol{p}\right\|^2 \tag{20}$$

式中，$\boldsymbol{P}$ 是预测网格表面的点，$\boldsymbol{Q}$ 是真实网格表面的点，第 1 项表示 $\boldsymbol{P}$ 中任意一点 $\boldsymbol{p}$ 到 $\boldsymbol{Q}$ 的最小距离之和，第 2 项则表示 $\boldsymbol{Q}$ 中任意一点 $\boldsymbol{q}$ 到 $\boldsymbol{P}$ 的最小距离之和。

2）法线一致性用于评估模型表面法向量与真值的一致性，通过比较两个点云的法向量相似性，评估它们之间的表面一致性，具体为

$$NC(\boldsymbol{P},\boldsymbol{Q}) = \frac{1}{|\boldsymbol{P}|}\sum_{p\in P}\max_{q\in Q}(\boldsymbol{n}_p\cdot\boldsymbol{n}_q) \tag{21}$$

对于点云 $\boldsymbol{P}$ 中的每个点 $\boldsymbol{p}$，找到点云 $\boldsymbol{Q}$ 中与其最近或法向量方向最接近的点 q，计算这两个点的法向量的点积，度量它们的法向一致性，对 $\boldsymbol{P}$ 中所有点的法向一致性进行平均，得到整体的法向一致性评分。

3）点到表面的距离是一种用于衡量点与表面之间距离的指标，考虑了表面的几何形状，能够更好地反映点与表面之间的真实距离。点 $\boldsymbol{p}$ 到表面 $\boldsymbol{S}$ 上最近点的欧几里得距离的计算式为

$$\boldsymbol{P2S}(\boldsymbol{p},\boldsymbol{S}) = \min_{s\in S}\left\|\boldsymbol{p}-\boldsymbol{s}\right\| \tag{22}$$

3.3 实验结果

本节对 PerRec 进行定量、定性分析以及消融实验，以展示模块的有效性。

3.3.1 定量结果

表 1 和表 2 展示了不同方法在测试集上的定量评估结果，包括倒角距离、点到面距离以及法线一致性 3 个关键指标。

通过与现有方法的对比分析，可以得出以下结论：PerRec 在所有评估指标上均展现出显著优势。

表 1 在 THuman2.0 数据集上的比较

Table 1 Comparison on the THuman2.0 dataset

方法	年份	倒角距离/cm ↓	法线一致性 ↑	点到表面距离/cm ↓
PIFu	2019	13.287	36.960	4.950
PIFuhd	2020	14.091	35.266	5.710
PaMIR	2022	28.885	31.203	5.612
ICON	2022	17.184	32.527	7.653
ECON	2023	17.963	30.295	7.855
SITH	2024	13.413	34.778	4.771
FOFX	2024	11.445	41.163	4.123
PerRec	2025	**1.957**	**66.700**	**0.983**

注：加粗字体表示各列最优结果。“↑”表示值越高越好，“↓”表示值越低越好。

表 2 在 CustomHumans 数据集上的比较

Table 2 Comparison on the CustomHumans dataset

方法	年份	倒角距离/cm ↓	法线一致性 ↑	点到表面距离/cm ↓
PIFu	2019	13.888	36.982	4.903
PIFuhd	2020	12.803	37.178	5.579
PaMIR	2022	30.404	29.994	4.857
ICON	2022	14.856	34.546	6.934
ECON	2023	20.039	29.912	9.472
SITH	2024	13.912	34.524	4.653
FOFX	2024	8.433	39.300	6.103
PerRec	2025	**2.336**	**72.603**	**0.919**

注：加粗字体表示各列最优结果。“↑”表示值越高越好，“↓”表示值越低越好。

具体而言，在倒角距离方面，相较于当前最优方法在 Thuman2.0（Yu 等，2021）和 CustomHumans（Ho 等，2023）数据集上分别提升 9.488 cm 和 6.097 cm。在点到表面距离指标上，PerRec 达到 0.983 cm 和 0.919 cm，较次优方法 4.123 cm 和 4.653 cm 分别提升 76.16% 和 80.25%。在法线一致性上，PerRec 在 Thuman2.0 和 CustomHumans 数据集上达到了 66.7 和 72.603，相较于当前最优方法提升了 62.04% 和 84.74%，这充分验证了所提方法在几何细节重建方面的有效性。这种显著的性能改进可以归因于：首先，引入的多尺度特征融合机制有效捕捉了局部几何细节；其次，设计的自适应采样策略提高了关键

区域的点云密度；再次，引入了三维空间信息，弥补了在二维空间中细节的缺失；最后，引入伪多视角，缓解了单幅图像造成的固有视角局限性。此外，通过消融实验进一步验证了各模块的贡献度。

3.3.2　定性结果

PerRec在Thuman2.0和CustomHumans数据集的测试效果如图3所示。可以看出，SITH重建较细的四肢末端，FOFX、PaMIR（Zheng等，2022）和PIFu会出现断裂的肢体与漂浮伪影，ECON重建出粘连的肢体，ICON的重建结果表面细节粗糙。相比之下，PerRec在宏观结构上和细节特征上均有较好的表现。重建出的人体四肢比例合理，有着较好的完整性。特别是在人体姿态纵向深度较大时，本文方法能展示明显的重建优势。

3.3.3　消融实验

为了验证各模块的有效性，在THuman2.0（Yu等，2021）和CustomHumans（Ho等，2023）数据集进行系统的消融实验。主要围绕以下3个模块展开：1）透视图像特征自适应模块，用于缓解图像中由相机投影引起的非均匀畸变；2）三维几何表征模块，通过傅里叶编码的SMPL占有率场建模三维几何连续性；3）伪多视角模块，通过虚拟光心移动策略增强模型对不同畸变区域的鲁棒性。

从无任何模块的基线模型出发，逐步引入上述模块，并观察性能变化。所有实验均采用相同的训练设置与网络结构，仅在模块组合上存在差异。结果如表3所示。可以看出，随着模块的逐步加入，模型在各项指标上均取得显著提升。

仅使用透视图像特征自适应模块时，模型能够有效校正不同图像区域的透视畸变。相比基线模型，倒角距离（CD）在THuman2.0和CustomHumans数据集上分别下降了62.1%与48.6%，法线一致性（NC）分别提升27.3%与28.2%，说明该模块显著改善了整体几何对齐精度。

图3　在THuman2.0和CustomHumans数据集的定性结果

Fig. 3　Qualitative results on the THuman2.0 and CustomHumans datasets

表 3　消融实验定量结果

Table 3　Ablation experiment quantitative results

方法			Thuman2.0			CustomHumans		
透视图像特征自适应	三维几何表征	伪多视角	倒角距离↓	法线一致性↑	P-to-S↓	倒角距离↓	法线一致性↑	P-to-S↓
–	–	–	12.837	46.312	4.864	12.058	45.874	4.518
√	–	–	4.862	58.958	2.522	6.201	58.832	3.348
–	√	–	3.316	65.264	2.038	4.308	68.331	1.902
√	√	–	2.221	**66.700**	1.030	2.360	71.635	0.946
√	√	√	**1.957**	**66.700**	**0.983**	**2.336**	**72.603**	**0.919**

注：加粗字体表示各列最优结果。“√”表示使用对应模块，“–”表示未使用。“↑”表示值越高越好，“↓”表示值越低越好。

仅使用三维几何表征时，模型在保持全局结构一致性的同时，显著增强了细节恢复能力。相较于基线，CD 分别下降 74.2% 与 64.3%，点到表面距离（P-to-S）下降 58.1% 与 57.9%，验证了傅里叶几何表征在捕捉高频形状特征方面的有效性。

同时启用透视图像特征自适应与三维几何表征后，性能进一步提升。在 THuman2.0 数据集上，CD 降至 2.221，NC 提升至 66.700，表明两者结合能够在全局结构与局部细节间取得更优平衡。

进一步引入伪多视角模块后，即完整的 PerRec 框架，模型在两个数据集的所有指标上均达到最优。伪多视角模块通过虚拟光心变换增强了模型在不同透视畸变分布下的鲁棒性，使重建在不同视角与人体姿态下保持一致性。需要说明的是，伪多视角模块本质上是一种数据增强策略，而非透视畸变校正的核心模块。在缺乏透视感知（即未结合透视特征自适应）的情况下，直接引入多视角可能引入虚假畸变并导致伪影，因此未单独对伪多视角模块进行消融。如图 4 所示，在不使用提出的模块时，模型无法恢复完整的三维人体结构，常出现四肢断裂与漂浮伪影。加入透视图像特征自适应模块后，整体形态得到修复，但四肢仍存在不连续。引入三维几何表征后，细节重建明显改善，而完整的 PerRec 在衣物褶皱、身体边缘等区域实现了最精细、最自然的重建效果。

3.3.4　采样点策略对比

在三维网格模型的隐式函数重建中，采样点的分布对模型的准确性具有重要影响。若在三维空间中随机采样，大多数采样点将远离网格表面，导致模型难以捕捉细节特征；反之，若采样点过于集中于网格表面，则可能导致过拟合现象，降低模型的泛化能力。故本模型的采样策略分为两个主要步骤：表

图 4　消融实验对比结果

Fig. 4　Comparative results of ablation experiments

面采样与空间采样。首先，从给定的三维网格模型中均匀采样表面点，并为每个采样点添加一定的随机噪声，以增加数据的多样性。这种噪声的引入有助于模型更好地学习表面的局部特征，避免过拟合。其次，在三维空间范围[-1, -1, -1]到[1]内随机生成额外的点，以覆盖模型的内部和外部空间。通过这种方式，确保采样点不仅分布在表面附近，还能充分覆盖模型的整体空间结构。为了确定每个采样点是否位于网格内部，本文采用多线程并行计算的方式进行批量处理。

在初步实验中，发现上述采样策略存在细节区域采样不足和粘连问题。对于身体部分较小（如手部）的区域，采样点数量不足，导致模型难以准确重建细节。在某些姿势下，由于胳膊和头部距离较近，模型无法对采样点进行分类，导致重建结果出现粘连现象。

为解决上述问题，本文进一步优化了采样策略，具体方法如下：利用SMPL模型的真值信息，提取手臂和手部的顶点，并让这些顶点沿着法线方向分别移动0.005、0.01、0.02等距离，生成额外的采样点。这种方法显著增加了手部区域的采样点数量，同时确保了体内外采样点的均匀分布。

表面采样与空间采样的结合能够有效提高模型的泛化能力，避免过拟合。手部采样增强策略显著改善了细节区域的重建精度，尤其是在手部和手臂部分。法线距离的选择对重建效果具有重要影响，本文通过表面采样、空间采样以及细节区域增强的有机结合，选择0.03作为法线移动距离，显著提升了三维网格模型隐式函数重建的准确性和鲁棒性。结果如图5所示。增加手部采样点策略的方案，有效地解决了细节区域采样点不足和粘连问题。

3.4 局限性

本文提出的畸变特征解耦与伪多视角约束的透视图像三维穿衣人体重建方法（PerRec）在处理透视畸变图像的穿衣人体重建任务上取得了最先进的水平，但是仍然存在一些问题，需要在今后的工作中进一步研究解决。

3.4.1 手部细节重建效果不佳

手部细节重建效果不佳的根源在于其复杂的几何结构。一方面，手部区域往往仅占图像面积的较小区域导致局部特征提取不充分；另一方面，手部频繁的较小区域，相互遮挡（如握拳、交叉手指）及透视畸变引发的关节比例失真，使得传统基于全身统一特征提取的方法难以捕捉指尖曲率、指甲形态等微几何特征。对此，受之前的工作启发（张冀 等，2025），可构建手部感知的重建架构——在全局人体重建网络基础上，引入轻量级手部局部优化模块，通过手部语义分割与关节热力图定位高关注区域。

图5 手部区域采样策略有效性验证

Fig. 5 Validation of hand region sampling strategy

3.4.2 遮挡场景

传统三维重建方法在遮挡情况下往往失效，主要原因是被遮挡区域缺乏几何信息，且单视角观测难以推断合理结构。为此，可利用Transformer的自注意力机制建立遮挡区域与可见部位的全局关联，然后融合虚拟多视角信息进行几何一致性约束，最后通过概率优化生成完整合理的重建结果。

4 结 论

本文针对透视畸变图像的穿衣人体重建展开研究。现有基于正交/弱透视投影的方法在近距离拍摄时存在非线性几何畸变和深度信息缺失问题，对此提出以下创新性解决方案。1）提出畸变特征解耦与伪多视角约束的透视图像三维穿衣人体重建方法，旨在解决透视图像穿衣人体重建任务；2）提出透视畸变特征自适应模块，解决传统透视投影机制造成的非线性几何形变；3）提出伪多视角模块，模拟多视角成像约束，有效缓解单图像信息缺失问题。相较现有方法，本文所提方法在Thuman2.0和CustomHumans数据集的定量和定性均达到最优，倒角距离、法线一致性和点到表面距离3个指标比目前最

优方法分别提高82.9%、62.0%、76.16%和72.3%、84.7%、80.2%。

参考文献(References)

Feng Q, Liu Y B, Lai Y K, Yang J Y and Li K. 2022. FOF: learning fourier occupancy field for monocular real-time human reconstruction//Proceedings of 2022 Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022. New Orleans, USA: [s.n.]: 7397-7409

Feng Q, Yang Y W, Liu Y B, Lai Y K, Yang J Y and Li K. 2024. FOF-X: towards real-time detailed human reconstruction from a single image [EB/OL]. [2025-09-05]. https://arxiv.org/pdf/2412.05961.pdf

Heikkila J and Silven O. 1997. A four-step camera calibration procedure with implicit image correction//Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Juan, USA: IEEE: 1106-1112 [DOI: 10.1109/cvpr.1997.609468]

Ho H I, Song J and Hilliges O. 2024. SiTH: single-view textured human reconstruction with image-conditioned diffusion//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 538-549 [DOI: 10.1109/cvpr52733.2024.00058]

Ho H I, Xue L X, Song J and Hilliges O. 2023. Learning locally editable virtual humans//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, USA: IEEE: 21024-21035 [DOI: 10.1109/cvpr52729.2023.02014]

Huang Q P, Liu L, Fu X D, Liu L J and Peng W. 2024. Clothed feature learning for single-view 3D human reconstruction. Journal of Image and Graphics, 29(9): 2610-2624 (黄千芃, 刘骊, 付晓东, 刘利军, 彭玮. 2024. 单视角三维人体重建的着装特征学习. 中国图象图形学报, 29(9): 2610-2624) [DOI: 10.11834/jig.230623]

Huang Z, Xu Y L, Lassner C, Li H and Tung T. 2020. Arch: animatable reconstruction of clothed humans//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 3090-3099 [DOI: 10.1109/cvpr42600.2020.00316]

Jiang W, Yi K M, Samei G, Tuzel O and Ranjan A. 2022. NeuMan: neural human radiance field from a single video//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 402-418 [DOI: 10.1007/978-3-031-19824-3_24]

Li Z H, Liu J Z, Zhang Z S, Xu S C and Yan Y L. 2022. CLIFF: carrying location information in full frames into human pose and shape estimation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 590-606 [DOI: 10.1007/978-3-031-20065-6_34]

Liu L Y, Sun J C, Gao Y Q, Gao C X and Chen J Y. 2024. Review of single-image 3D human reconstruction based on deep learning. Journal of Huazhong University of Science and Technology (Natural Science Edition), 52(5): 98-122 (刘乐元, 孙见弛, 高韵琪, 高常鑫, 陈靓影. 2024. 基于深度学习的单图像三维人体重建研究综述. 华中科技大学学报(自然科学版), 52(5): 98-122) [DOI: 10.13245/j.hust.240614]

Loper M, Mahmood N, Romero J, Pons-Moll G and Black M J. 2015. SMPL: a skinned multi-person linear model. ACM Transactions on Graphics, 34(6): #248 [DOI: 10.1145/2816795.2818013]

Osman A A A, Bolkart T and Black M J. 2020. STAR: sparse trained articulated human body regressor//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 598-613 [DOI: 10.1007/978-3-030-58539-6_36]

Pavlakos G, Choutas V, Ghorbani N, Bolkart T, Osman A A, Tzionas D, et al. 2019. Expressive body capture: 3D hands, face, and body from a single image//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 10967-10977 [DOI: 10.1109/cvpr.2019.01123]

Pu J C, Liu L, Fu X D, Liu L J and Huang Q S. 2022. Clothing visual representation for 3D human reconstruction. Journal of Computer-Aided Design and Computer Graphics, 34(3): 352-363 (普骏程, 刘骊, 付晓东, 刘利军, 黄青松. 2022. 三维人体重建中的服装视觉信息表示. 计算机辅助设计与图形学学报, 34(3): 352-363) [DOI: 10.3724/SP.J.1089.2022.18921]

Saito S, Huang Z, Natsume R, Morishima S, Li H and Kanazawa A. 2019. PIFu: pixel-aligned implicit function for high-resolution clothed human digitization//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 2304-2314 [DOI: 10.1109/iccv.2019.00239]

Saito S, Simon T, Saragih J and Joo H. 2020. PIFuHD: multi-level pixel-aligned implicit function for high-resolution 3D human digitization//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 81-90 [DOI: 10.1109/cvpr42600.2020.00016]

Saito S, Yang J L, Ma Q L and Black M J. 2021. SCANimate: weakly supervised learning of skinned clothed avatar networks//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 2885-2896 [DOI: 10.1109/cvpr46437.2021.00291]

Tang Y Z, Zhang Q J and Hou J H. 2026. HuGDiffusion: generalizable single-image human rendering via 3D Gaussian diffusion. IEEE Transactions on Visualization and Computer Graphics, 32(2): 2061-2074 [DOI: 10.1109/TVCG.2025.3625230]

Wang W J, Ge Y T, Mei H Y, Cai Z G, Sun Q P, Wang Y J, et al. 2023. Zolly: zoom focal length correctly for perspective-distorted human mesh reconstruction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France:

IEEE: 3902-3912 [DOI: 10.1109/iccv51070.2023.00363]

Xiu Y L, Yang J L, Cao X, Tzionas D and Black M J. 2023. ECON: explicit clothed humans optimized via normal integration//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 512-523 [DOI: 10.1109/cvpr52729.2023.00057]

Xiu Y L, Yang J L, Tzionas D and Black M J. 2022. ICON: implicit clothed humans obtained from normals//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 13286-13296 [DOI: 10.1109/cvpr52688.2022.01294]

Yu T, Zheng Z R, Guo K W, Liu P P, Dai Q H and Liu Y B. 2021. Function4D: real-time human volumetric capture from very sparse consumer RGBD sensors//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 5742-5752 [DOI: 10.1109/cvpr46437.2021.00569]

Zhang J, Ren Z P, Zhang R H, Yuan C, Zhai Y J and Yu Z Q. 2025. Hand-focused reconstruction of monocular RGB clothed humans. Application Research of Computers, 42(1): 300-306 (张冀, 任志鹏, 张荣华, 苑朝, 翟永杰, 余正秦. 2025. 单目RGB穿衣人体的手部精细化重建. 计算机应用研究, 42(1): 300-306) [DOI: 10.19734/j.issn.1001-3695.2024.02.0112]

Zhang Z. 2000. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11): 1330-1334 [DOI: 10.1109/34.888718]

Zheng Z R, Yu T, Liu Y B and Dai Q H. 2022. PaMIR: parametric model-conditioned implicit representation for image-based human reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6): 3170-3184 [DOI: 10.1109/tpami.2021.3050505]

作者简介

杨源旺,男,硕士研究生,主要研究方向为计算机视觉和计算机图形学。E-mail: yyw@tju.edu.cn

李坤,通信作者,女,教授,博士生导师,主要研究方向为计算机视觉与图形学、人工智能和多媒体处理。

E-mail: lik@tju.edu.cn

马健,通信作者,男,助理研究员,硕士生导师,主要研究方向为三维视觉。E-mail:jianma@tju.edu.cn

孟源,女,硕士研究生,主要研究方向为计算机视觉和计算机图形学。E-mail: ghqb480736_@tju.edu.cn